

# Exponential Order Statistics and Some Combinatorial Identities

P. Vellaisamy

Department of Mathematics  
Indian Institute of Technology Bombay  
Powai, Mumbai-400076, India.  
Email: *pv@math.iitb.ac.in*

## Abstract

We consider the  $k$ -th order statistic from unit exponential distribution and, by computing its Laplace transform, show that it can be represented as sum of independent exponential rvs. Our proof is simple and different. It readily proves that the standardized exponential spacings also follow unit exponential distribution. Another advantage of our approach is that by computing the Laplace transform of the  $k$ -th order statistic in two different ways, we derive several interesting combinatorial identities. A probabilistic interpretation of these identities and their generalizations are also given.

**Keywords.** Binomial inversion, combinatorial identities, exponential distribution, Laplace transform, order statistics, probabilistic proofs.

## 1 Introduction

It is known that order statistics from exponential distribution have several interesting properties. We consider, without loss of generality, the exponential distribution with mean unity. For example, the  $k$ -th order statistic has the distribution of sum of independent exponential random variables (rvs). Another interesting result is that the spacings of order statistics also follow exponential distribution. The usual proofs of these results use the transformation to the set of spacings from the set of order statistics and by applying Jacobian density theorem.

In this paper, we prove the above-mentioned and some other results using the Laplace transform methods. This approach is simpler and indeed several related results can be proved in a unified way. Another key purpose of this article is to bring out the connection between exponential order statistics and several combinatorial identities. In fact, we give simpler proofs

of several combinatorial/binomial identities by evaluating the Laplace transformation of the  $k$ -th exponential order statistic by two different ways and equating them. We also point out the probabilistic interpretations of these combinatorial identities. The approach followed in the paper is similar to Vellaisamy (2015), where only the largest order statistic is considered.

## 2 The Order Statistics Results

Let  $X_1, \dots, X_n$  be iid continuous random variables (rvs) with cdf  $F(x)$  and density  $f(x)$ . Also, let  $X_{(1)} < X_{(2)} < \dots < X_{(n)}$  be the corresponding order statistics. Then it is well known (see David and Nagaraja (2003)) that the distribution and the density of  $k$ -th order statistic  $X_{(k)}$  are respectively given by

$$F_k(x) = P(X_{(k)} < x) = \sum_{m=k}^n \binom{n}{m} F^m(x) (1 - F(x))^{n-m}, \quad (2.1)$$

and

$$f_k(x) = n \sum_{m=k}^n \binom{n-1}{k-1} F^{k-1}(x) (1 - F(x))^{n-k} f(x), \quad (2.2)$$

for  $1 \leq k \leq n$ .

Our interest is on the exponential order statistics. Let  $T_1, \dots, T_n$  be independent unit exponential random variables (rvs) so that

$$f_{T_1}(t) = e^{-t}, \quad t > 0. \quad (2.3)$$

and let  $T_{(1)} < T_{(2)} < \dots < T_{(n)}$  be the order statistics of  $T_j$ 's. Then the density of  $T_{(k)}$  is

$$f_k(t) = \frac{1}{B(k, n-k+1)} (1 - e^{-t})^{k-1} e^{-(n-k+1)t}, \quad t > 0. \quad (2.4)$$

First we prove an interesting property of  $T_{(k)}$  by computing the Laplace transform of  $T_{(k)}$ .

Note for  $s > 0$ ,

$$\mathbb{E}(e^{-sT_{(k)}}) = \frac{1}{B(k, n-k+1)} \int_0^\infty e^{-(s+1)t} (1 - e^{-t})^{k-1} e^{-(n-k)t} dt.$$

Substitute  $w = (1 - e^{-t})$  to obtain

$$\begin{aligned} \mathbb{E}(e^{-sT_{(k)}}) &= \frac{1}{B(k, n-k+1)} \int_0^1 w^{k-1} (1-w)^{n-k+s+1-1} dw \\ &= \frac{B(k, n-k+s+1)}{B(k, n-k+1)} \\ &= \frac{\Gamma(n+1)\Gamma(n-k+s+1)}{\Gamma(n-k+1)\Gamma(n+s+1)}. \end{aligned}$$

Using

$$\begin{aligned}\Gamma(n+1) &= n(n-1)\dots(n-k+1)\Gamma(n-k+1); \\ \Gamma(s+1+n) &= (s+1+n-1)\dots(s+1+n-k)\Gamma(s+1+n-k),\end{aligned}$$

we obtain

$$\begin{aligned}\mathbb{E}(e^{-sT_{(k)}}) &= \frac{n(n-1)(n-2)\dots(n-k+1)}{(s+n)(s+n-1)\dots(s+n-k+1)} \\ &= \prod_{j=n-k+1}^n \left( \frac{j}{s+j} \right) \\ &= f_{n,k}(s) \quad (\text{say}).\end{aligned}\tag{2.5}$$

Note also that

$$\begin{aligned}\mathbb{E}(e^{-sT_{(k)}}) &= \prod_{j=n-k+1}^n \mathbb{E}(e^{-sY_j}) \\ &= \prod_{j=n-k+1}^n \mathbb{E}(e^{-s \sum_{j=n-k+1}^n Y_j}),\end{aligned}\tag{2.6}$$

where  $Y_1, \dots, Y_n$  are independent exponential rvs and  $Y_j \sim \text{Exp}(j)$  with mean  $j^{-1}$ ,  $1 \leq j \leq n$ . Thus, for  $1 \leq k \leq n$ , we have from (2.6)

$$T_{(k)} \stackrel{d}{=} \sum_{j=n-k+1}^n Y_j,\tag{2.7}$$

where  $X \stackrel{d}{=} Y$  means both  $X$  and  $Y$  have identical distributions.

Note when  $k = n$ , we have

$$T_{(n)} \stackrel{d}{=} \sum_{j=1}^n Y_j,$$

a known result (see, for example, (4.1) of Vellaisamy (2015)).

Also, when  $k = 1$ ,

$$T_{(1)} \stackrel{d}{=} Y_n,$$

where  $T_n \sim \text{Exp}(n)$ , a well-known result.

Next some interesting remarks are in order.

**Remark 2.1** From (2.7),

$$\begin{aligned}T_{(k)} &\stackrel{d}{=} Y_{n-k+1} + Y_{n-k+2} + \dots + Y_n \\ &\stackrel{d}{=} W_1 + W_2 + \dots + W_k,\end{aligned}\tag{2.8}$$

where  $W_j$ 's are independent and  $W_j \sim \text{Exp}(n - k + j)$ . Also, from (2.8), we get

$$\mathbb{E}(T_{(k)}) = \sum_{j=1}^k \mathbb{E}(W_j) = \sum_{j=1}^k \frac{1}{(n - k + j)} \quad (2.9)$$

and

$$\mathbb{V}ar(T_{(k)}) = \sum_{j=1}^k \mathbb{V}ar(W_j) = \sum_{j=1}^k \frac{1}{(n - k + j)^2}, \quad (2.10)$$

for  $1 \leq k \leq n$ .

Note that usual proofs of the results in (2.9) and (2.9), based on the density of  $T_{(k)}$ , are rather complicated.

**Remark 2.2** Also, from the representation given in (2.7), we can immediately obtain the distribution of the spacings  $(T_{(k)} - T_{(k-1)})$ ,  $1 \leq k \leq n$ , with  $T_{(0)} \equiv 0$ .

Since

$$T_{(k)} \stackrel{d}{=} Y_{n-k+1} + Y_{n-k+2} + \dots + Y_n$$

we have for  $1 \leq k \leq n$ ,

$$T_{(k)} - T_{(k-1)} \stackrel{d}{=} Y_{n-k+1} \stackrel{d}{=} \frac{T_1}{(n - k + 1)}$$

or equivalently the normalized spacings

$$(n - k + 1)(T_{(k)} - T_{(k-1)}) \stackrel{d}{=} T_1, \quad 1 \leq k \leq n, \quad (2.11)$$

where  $T_1 \sim \text{Exp}(1)$ .

The usual proof of the above result is based on the joint density of  $T_{(1)}, \dots, T_{(k)}$  and then making suitable transformations to spacings. This is rather involved, whereas our proof easily follows from (2.7).

**Remark 2.3** Let  $Z_n = (T_{(n)} - \ln(n))$ . Since,

$$P(Z_n \leq x) = (1 - e^{-(\ln(n)+x)})^n = (1 - \frac{e^{-x}}{n})^n \longrightarrow \exp\{-e^{-x}\}, \text{ as } n \rightarrow \infty,$$

we have  $Z_n \xrightarrow{\mathcal{L}} Z$  (see Billingsley (1995, p. 329) for the definition of convergence in distribution), where  $Z$  follows standard Gumbel distribution with density

$$f(z) = e^{-z} \exp(-e^{-z}), \quad z \in \mathbb{R}.$$

It is known that  $E(Z) = \nu$  (Euler's constant) and  $Var(Z) = \frac{\pi^2}{6}$ .

Hence, we have

$$\lim_{n \rightarrow \infty} E(Z_n) = \lim_{n \rightarrow \infty} \left( \sum_{j=1}^n \frac{1}{j} - \ln(n) \right) = \nu. \quad (2.12)$$

and

$$\lim_{n \rightarrow \infty} Var(Z_n) = \lim_{n \rightarrow \infty} \left( \sum_{j=1}^n \frac{1}{j^2} \right) = \frac{\pi^2}{6}. \quad (2.13)$$

Note that other mathematical proofs for the equality of the last two terms of (2.12) and (2.13) are rather involved. More importantly, it is interesting to observe the connections between the exponential order statistics and the above results.

### 3 Combinatorial Identities

In this section, we derive several combinatorial identities using the Laplace transformation of the  $k$ -th order statistic  $T_{(k)}$ . The basic idea, from Vellaisamy (2015), is to derive the Laplace transform in two different ways and equate them to get identities. Note that we have already derived one form, using the pdf of  $T_{(k)}$ . We obtain next another form using the cdf of  $T_{(k)}$  and the integration by parts.

First note that the cdf of  $T_{(k)}$  from (2.1) is

$$F_k(x) = \sum_{m=k}^n \binom{n}{m} (1 - e^{-t})^m e^{-(n-m)t}, \quad t > 0. \quad (3.1)$$

Using (3.1), we have for  $1 \leq k \leq n$ ,

$$\begin{aligned} \mathbb{E}(e^{-sT_{(k)}}) &= \int_0^\infty e^{-st} dF_k(t) \\ &= s \int_0^\infty e^{-st} F_k(t) dt \quad (\text{integration by parts}) \\ &= s \int_0^\infty e^{-st} \left\{ \sum_{m=k}^n \binom{n}{m} (1 - e^{-t})^m e^{-(n-m)t} \right\} dt \\ &= s \sum_{m=k}^n \binom{n}{m} \int_0^\infty e^{-st} \left\{ \sum_{j=0}^m \binom{m}{j} (-1)^j e^{-jt} \right\} e^{-(n-m)t} dt \\ &= s \sum_{m=k}^n \binom{n}{m} \sum_{j=0}^m (-1)^j \binom{m}{j} \int_0^\infty e^{-(s+j+n-m)t} dt \end{aligned}$$

$$\begin{aligned}
&= \sum_{m=k}^n \sum_{j=0}^m (-1)^j \binom{n}{m} \binom{m}{j} \left( \frac{s}{s+j+(n-m)} \right) \\
&= g_{n,k}(s), \quad (\text{say}).
\end{aligned} \tag{3.2}$$

Thus, from (2.5) and (3.2), we obtain an interesting combinatorial identity

$$\sum_{m=k}^n \sum_{j=0}^m (-1)^j \binom{n}{m} \binom{m}{j} \left( \frac{s}{s+n-m+j} \right) = \prod_{j=n-k+1}^n \left( \frac{j}{s+j} \right), \tag{3.3}$$

for  $1 \leq k \leq n$ , and  $s > 0$ .

When  $k = 1$ , we get from (3.3),

$$\sum_{m=1}^n \sum_{j=0}^m (-1)^j \binom{n}{m} \binom{m}{j} \left( \frac{s}{s+n-m+j} \right) = \frac{n}{s+n}, \tag{3.4}$$

for  $n \geq 1$  and  $s > 0$ , an interesting binomial identity.

Similarly when  $k = n$ , we obtain from (3.3),

$$\sum_{j=0}^n (-1)^j \binom{n}{j} \left( \frac{s}{s+j} \right) = \prod_{j=1}^n \left( \frac{j}{s+j} \right). \tag{3.5}$$

**Remark 3.1** Recently, Peterson (2013) derived the above binomial identity from probabilistic considerations. Vellaisamy (2015) gave a simple and different proof, based on Laplace transforms. He also discussed a statistical application of the above identity, using the binomial inversion formula (see Aigner (2007), p. 73). Note also that an application of binomial inversion to (3.5) leads to equation (4.7) of Vellaisamy (2015). Similarly, the equation (3.3) will lead to a new identity.

**Remark 3.2** Note from (3.3),

$$\sum_{m=k}^n \binom{n}{m} \left( \frac{s}{s+n-m} \right) \left\{ \sum_{j=0}^m (-1)^j \binom{m}{j} \left( \frac{s+n-m}{s+n-m+j} \right) \right\} = \prod_{j=n-k+1}^n \left( \frac{j}{s+j} \right),$$

which using (3.5), leads to

$$\sum_{m=k}^n \binom{n}{m} \left( \frac{s}{s+n-m} \right) \prod_{i=1}^m \left( \frac{i}{s+n-m+i} \right) = \prod_{j=n-k+1}^n \left( \frac{j}{s+j} \right),$$

for  $1 \leq k \leq n$ , and  $s > 0$ .

## 4 Probabilistic Connections and Generalizations

In this section, we give a probabilistic interpretation to the main combinatorial identity

$$f_{n,k}(s) = g_{n,k}(s)$$

for  $n \geq 1$  and  $s > 0$ .

### 4.1 Probabilistic Interpretations

Let now  $X_r \sim G(s, r)$ ,  $s > 0$ ,  $r \geq 1$  be a positive integer, with density

$$f(x|s, r) = \frac{s^r}{\Gamma(r)} e^{-sx} x^{r-1}, \quad x > 0. \quad (4.1)$$

Then it is well-known (see eq. (3.3.9) of Casella and Berger (2002)) that

$$\mathbb{P}(X_r > x) = \sum_{j=0}^{r-1} \frac{e^{-sx} (sx)^j}{j!}. \quad (4.2)$$

Using (4.2), we obtain

$$\begin{aligned} \mathbb{P}(X_r > T_{(k)}) &= \mathbb{E} [\mathbb{P}(X_r > T_{(k)}) | T_{(k)}] \\ &= \sum_{j=0}^{r-1} \frac{s^j}{j!} \mathbb{E} \left( e^{-sT_{(k)}} T_{(k)}^j \right). \end{aligned} \quad (4.3)$$

Note that  $f_{n,k}(s) = \mathbb{E}(e^{-sT_{(k)}})$  satisfies

$$f_{n,k}^{(j)}(s) = (-1)^j \mathbb{E} \left( e^{-sT_{(k)}} T_{(k)}^j \right), \quad (4.4)$$

where  $h^{(j)}(s)$  denotes the  $j$ -th derivative of  $h(s) = h^{(0)}(s)$ .

Thus, for  $r \geq 1$  and  $s > 0$ , we obtain from (4.3) and (4.4),

$$\mathbb{P}(X_r > T_{(k)}) = \sum_{j=0}^{r-1} (-1)^j \frac{s^j}{j!} f_{n,k}^{(j)}(s) \quad (4.5)$$

$$= \sum_{j=0}^{r-1} (-1)^j \frac{s^j}{j!} g_{n,k}^{(j)}(s), \quad (4.6)$$

since  $f_{n,k}(s) = g_{n,k}(s)$ .

When  $r = 1$ , we obtain

$$f_{n,k}(s) = g_{n,k}(s) = P(X_1 > T_{(k)}), \quad (4.7)$$

which shows that the binomial identity (3.3) admits a probabilistic interpretation.

Similarly, when  $r = 2$ , we get

$$f_{n,k}(s) - sf'_{n,k}(s) = g_{n,k}(s) - sg'_{n,k}(s) \quad (4.8)$$

which is a new identity (see (4.13)) and this corresponds to  $P(X_2 > T_k)$ .

Note however the combinatorial identity obtained from  $f'_{n,k}(s) = g'_{n,k}(s)$  may not represent a probability of an event (see Vellaisamy (2015), p. 243).

## 4.2 Generalizations

In this section, we generalize some of the binomial identities. Observe that

$$\begin{aligned} \mathbb{P}(T_{(k)} < X_r) &= \mathbb{E} [\mathbb{P}(T_{(k)} < X_r) | X_r] \\ &= \mathbb{E} \left( \sum_{m=k}^n \binom{n}{m} (1 - e^{-X_r})^m e^{-(n-m)X_r} \right) \quad (\text{from (3.1)}) \\ &= \sum_{m=k}^n \binom{n}{m} \left( \sum_{j=0}^m (-1)^j \binom{m}{j} \mathbb{E} e^{-(n-m+j)X_r} \right) \end{aligned} \quad (4.9)$$

Since  $X_r \sim G(s, r)$ , we have

$$\mathbb{E} e^{-(n-m+j)X_r} = \left( \frac{s}{s + n - m + j} \right)^r. \quad (4.10)$$

Hence, we get

$$\mathbb{P}(T_{(k)} < X_r) = \sum_{m=k}^n \binom{n}{m} \sum_{j=0}^m (-1)^j \binom{m}{j} \left( \frac{s}{s + n - m + j} \right)^r. \quad (4.11)$$

Thus, we obtain from (4.5) and (4.11),

$$\sum_{m=k}^n \binom{n}{m} \sum_{j=0}^m (-1)^j \binom{m}{j} \left( \frac{s}{s + n - m + j} \right)^r = \sum_{j=0}^{r-1} (-1)^j \frac{s^j}{j!} f_{n,k}^{(j)}(s) \quad (4.12)$$

where as seen before

$$f_{n,k}(s) = \prod_{j=n-k+1}^n \left( \frac{j}{s+j} \right).$$

When  $r = 1$ , the identity in (4.12) reduces to (3.3).



When  $r = 2$ , we get

$$\begin{aligned} \sum_{m=k}^n \binom{n}{m} \sum_{j=0}^m (-1)^j \binom{m}{j} \left( \frac{s}{s+n-m+j} \right)^2 &= f_{n,k}(s) - s f_{n,k}^{(1)}(s) \\ &= \prod_{j=n-k+1}^n \left( \frac{j}{s+j} \right) \left[ 1 + \sum_{j=n-k+1}^n \frac{s}{s+j} \right] \end{aligned} \quad (4.13)$$

When  $k = n$ , the above identity reduces to equation (4.3) of Vellaisamy (2015). However, when  $k = 1$ , the above result gives us a new identity, namely,

$$\sum_{m=1}^n \binom{n}{m} \sum_{j=0}^m (-1)^j \binom{m}{j} \left( \frac{s}{s+n-m+j} \right)^2 = \frac{n(n+2s)}{(s+n)^2}, \quad (4.14)$$

for  $s > 0$  and  $n \geq 1$ .

Thus, we have generalized the basic binomial identity in (3.5) in several directions. We hope the numerous identities derived in this paper could be useful to other applied areas as well.

## References

- Aigner, M. (2007). *A Course in Enumeration*. Berlin, Springer-Verlag.
- Billingsley, P. (1995). *Probability and Measure*. Third edition, John Wiley & Sons, New York.
- David, H. A. and Nagaraja, H. N. (2003). *Order Statistics*. Third edition, Wiley, New York.
- Casella, G. and Berger, R. L. (2002). *Statistical Inference*. Second edition, the Wadsworth & Brooks, CA.
- Peterson, J. (2013). A probabilistic proof of a binomial identity. *American Mathematical Monthly*, 120, 558-562.
- Vellaisamy, P. (2015). On probabilistic proofs of certain binomial identities. *American Statistician*, 69, 241-243.